END
DATE
FILMED
6-81
DTIC

LEVEL

MRC Technical Summary Report #2180 ✓

SOLVING QUADRATIC PROGRAMS BY AN EXACT
PENALTY FUNCTION

Shih-Ping Han

AD A099368

Mathematics Research Center
University of Wisconsin—Madison
610 Walnut Street
Madison, Wisconsin 53706

February 1981

(Received September 24, 1980)

DTIC
SELECTED
MAY 27 1981
A

Approved for public release
Distribution unlimited

X 81 5 27 011

UNIVERSITY OF WISCONSIN-MADISON
MATHEMATICS RESEARCH CENTER

SOLVING QUADRATIC PROGRAMS BY AN EXACT PENALTY FUNCTION

Shih-Ping Han[†]

Technical Summary Report #2180
February 1981

ABSTRACT

In this paper we study a gradient projection method for quadratic programs that does not require the generated points to be feasible and can avoid the computation of a feasible starting point. This is done by using an exact penalty function in the line-search. It is shown that the method can produce from any starting point a solution in a finite number of iterations.

AMS (MOS) Subject Classifications: 90C20, 90C25

Key Words: Quadratic programming, gradient projection method, exact penalty function

Work Unit Number 5 (Operations Research)

[†]Department of Mathematics, University of Illinois-Urbana, Urbana, Illinois, 61801

SIGNIFICANCE AND EXPLANATION

Many nonlinear optimization problems, such as general
nonlinear programming and minimax problems, can be tackled by
solving a sequence of quadratic programs. Therefore, it is
very important to develop efficient computational method for
quadratic programs. For such problems, we propose a gradient-
projection-type method that does not require that the generated
points be feasible. By so doing, we can avoid the expense of
computating a starting feasible point. The method is shown to
attain a solution in a finite number of steps and this has led
to satisfactory computational results.

Accession For

NTIS GRA&I
DTIC TAB
Unannounced
Justification

Distribution/
Availability Codes
Avail and/or
Special

A

The responsibility for the wording and views expressed in this
descriptive summary lies with MRC, and not with the author of
this report.

SOLVING QUADRATIC PROGRAMS BY AN EXACT PENALTY FUNCTION

Shih-Ping Han[†]

1. INTRODUCTION

We are concerned with the following quadratic programming problem

$$\min \quad \tfrac{1}{2}x^T M x - c^T x$$
$$\text{s.t.} \quad A x \leq b, \tag{QP}$$

where M and A are $n \times n$ and $m \times n$ matrices respectively. For convenience, we consider here only inequality constraints. The extension of our results to problems with equality constraints is apparent and has no major difficulty. We also assume that the matrix M is positive definite. However, for some type of semidefinite matrices, the theory in this paper can also go through and the proposed method is applicable.

Among the methods for quadratic programming, the class of gradient projection methods have been widely used and studied [1,2,3,6,7]. The basic steps of this type of methods can be described as follows. At a feasible point x, we find the index set $J = \{j \,|\, a_j^T x = b_j\}$ of active constraints at x and find the point $x^{(J)}$ that solves the following equality constrained quadratic programming problem

[†]Department of Mathematics, University of Illinois-Urbana, Urbana, Illinois, 61801

$$\min_{z} \quad \tfrac{1}{2}z^T M z - c^T z$$

$$\text{s.t.} \quad A_J z = b_J ,$$

(1.1)

where A_J and b_J are the submatrix and the subvector of A and b corresponding to the index set J respectively. A search direction d is then determined by

$$d = x^{(J)} - x .$$

The direction d can also be obtained by projecting the Newton direction $-M^{-1}\nabla f(x)$ onto the plane $\{z \mid A_J z = b_J\}$. More specifically, we have

$$d = -PM^{-1}\nabla f(x)$$

where $P = I - M^{-1}A_J^T(A_J M^{-1} A_J^T)^{-1}A_J$ is the projection matrix with weight M.

Once the direction d is chosen, a new point $\bar{x} = x + \bar{\lambda}d$ is computed by taking a stepsize $\bar{\lambda}$ that is the largest number in $[0,1]$ keeping $x + \lambda d$ feasible. This stepsize procedure can be viewed as a special case of the following scheme:

$$P(x + \bar{\lambda}d, \alpha) = \min_{\lambda \geq 0} P(x + \lambda d, \alpha) ,$$

where P is any exact penalty function studied by Han and Mangasarian [4] and defined as

$$P(x,\alpha) := \tfrac{1}{2}x^T M x - c^T x + \alpha \| (Ax-b)_+ \| .$$

(1.2)

Here, $\| \cdot \|$ is an arbitrary but fixed norm in R^m and $(Ax-b)_+$ is an m-vector with $\max\{a_i^T x - b_i, 0\}$ as its i-th component. If we prefer to maintain feasibility of our generated points then we may set the penalty parameter α

the value infinity. The resulting procedure is reduced to the one used in a usual gradient projection method.

To maintain feasibility, we need a starting feasible point, which is usually computed by a linear programming technique. Because the degree of complexity of linear programming is about equivalent to that of quadratic programming, the effort in finding a starting point is unproportionally large. Moreover, because the objective function is completely ignored in the process, the obtained starting point is often far from our solution. Furthermore, we sometimes have a good but infeasible estimate of a solution, as often in the case that we solve a sequence of quadratic programs to tackle a general nonlinear programming problem. It is difficult to fully exploit this useful information if we insist on maintaining feasibility.

In this paper we study a gradient projection method that allows the penalty parameter α to have a finite value in the computation of stepsizes. By doing so the feasibility requirement for the generated points is relaxed and a starting point no longer needs to be feasible. This approach can also be viewed as taking the objective function into consideration in the process of finding a feasible starting point and combining the usual two phases in a gradient projection method into one.

We also incorporate into this method a procedure in selecting active constraints so that the generated points may not always move along the edges of the feasible region. If the original problem is feasible and the gradients of active

constraints at any feasible point are linearly independent, then it is shown that the method can find a solution in a finite number of steps. Computational results show that the number of steps is almost always less than the number of constraints.

We describe the method in Section 2 and study its basic properties in Section 3. Convergence theorems are given in Section 4. Section 5 contains some computational results.

2. THE METHOD

To facilitate our description of the method we first introduce some notation. Given a point x, we use $J_+(x)$ to denote the index set $\{j \mid a_j^T x > b_j\}$ and use $J_0(x)$ to denote $\{j \mid a_j^T x = b_j\}$ and $J(x) = J_+(x) \cup J_0(x)$. Notice that $J(x)$ is just the index set of active constraints at x. When the reference of the point x is clear from context, we may drop the symbol x from $J_+(x)$, $J_0(x)$ and $J(x)$ and simply use J_+, J_0 and J, respectively.

Given an index set $I \subset \{1,2,\ldots,m\}$, we are interested in a solution of the following equality constrained problem

$$\min_{x} \quad \tfrac{1}{2}x^T M x - c^T x$$

$$\text{s.t.} \quad A_I x = b_I .$$

However, the above problem may be infeasible for some index set I. Therefore, we prefer to consider, instead, the following problem

$$\min_{x} \quad \tfrac{1}{2}x^T M x - c^T x \tag{2.1}$$

$$\text{s.t.} \quad A_I^T A_I x = A_I^T b_I .$$

Here, the minimization is over the set of all least-squares solutions of the system $A_I x = b_I$ and the problem is always feasible. In practice one would not compute the $n \times n$ matrix $A_I^T A_I$, but instead would make use of a QR decomposition of A_I. If the matrix M is positive definite, as is assumed in this paper, there exists a unique solution to (2.1). We denote this solution by $x^{(I)}$. If $v^{(I)}$ is a Lagrange multiplier vector of $x^{(I)}$, we define an m-vector

$u^{(I)}$ by $u_I^{(I)} := A_I v^{(I)}$ and $u_j^{(I)} := 0$ for $j \notin I$. Clearly, the pair $(x^{(I)}, u^{(I)})$ satisfy the following conditions:

(a) $Mx^{(I)} - c + A_I^T u_I^{(I)} = 0$

(b) $A_I^T A_I x^{(I)} = A_I^T b_I$ (2.2)

(c) $u_j^{(I)} = 0$ for $j \notin I$.

We assume that the multiplier vector $v^{(I)}$ is chosen in a consistent way when there are more than one such vectors. By so doing the vector $u^{(I)}$ is also uniquely determined by the index set I.

From a result in [4], for any vector norm $\|\cdot\|$ in R^m and for a sufficiently large number α, a solution of quadratic program (QP) is also a minimum point of the exact penalty function P defined in (1.2). In this paper we use the penalty function with the 2-norm, because it is compatible with least-squares solutions of the system $A_I x = b_I$.

In the method we improve our estimate x of a solution by finding a descent direction d for the exact penalty function P and search along this direction for a new point with a lower penalty function value. To obtain such a direction d, we consider the index set $J = J(x)$ and compute the pair $(x^{(J)}, u^{(J)})$. If $(x^{(J)}, u^{(J)})$ is not a Karush-Kuhn-Tucker pair of the original problem (QP) but $u_i^{(J)} \geq 0$ for all $i \in J_0(x)$, then we set $d = x^{(J)} - x$. If $u_i^{(J)} < 0$ for some $i \in J_0(x)$, we drop one of the constraints with a negative multiplier, say $u_j^{(J)}$, and set $I = J \setminus \{j\}$. Compute $(x^{(I)}, u^{(I)})$ for this new index set I and set $d = x^{(I)} - x$. We repeat this deleting process until either

$x^{(I)} \neq x^{(J)}$ or $u_i^{(I)} \geq 0$ for all $i \in J_0(x) \cap I$. It is noted here that if x is feasible and the vectors $\{a_i\}$, $i \in J(x)$, are linearly independent, then no more than one constraint can be deleted.

It is observed that the direction $x^{(I)} - x$ is not always better than $x^{(J)} - x$. Therefore, we set $d = x^{(J)} - x$ instead of $d = x^{(I)} - x$ when x is feasible, $x^{(J)} \neq x$ and

$$\tfrac{1}{2}(x^{(J)} - x^{(I)})^T M (x^{(J)} - x^{(I)}) < \alpha \| (Ax^{(I)} - b) \|_2 . \qquad (2.3)$$

This condition is equivalent to that

$$P(x^{(I)}, \alpha) > \tfrac{1}{2}(x^{(J)})^T M x^{(J)} - c^T x^{(J)} , \qquad (2.4)$$

and it is proved later that it helps the algorithm to terminate.

We sum up the procedure for generating the direction d at a point x as follows:

1. Find $J_+ = \{j \,|\, a_i^T x > b_i\}$,

   $J_0 = \{j \,|\, a_i^T x = b_i\}$,

   $J = J_+ \cup J_0$.

2. Set $I = J$.

3. Compute $(x^{(I)}, u^{(I)})$. If it is a Karush-Kuhn-Tucker pair for (QP), then stop.

4. If $x^{(I)} \neq x^{(J)}$ then go to 7.

5. If $u_j^{(I)} \geq 0$ for all $j \in J_0 \cap I$ then go to 7.

6. Find an index j such that

$$u_j^{(I)} = \min \{u_i^{(I)} | u_i^{(I)} < 0, \ i \in J_0 \cap I\}$$

and set $I = I \backslash \{j\}$. Go to 3.

7. If x is feasible, $x^{(J)} \neq x$ and condition (2.3) is satisfied, then set $d = x^{(J)} - x$; otherwise, $d = x^{(I)} - x$.

In the sequel, we use I to denote the index set that produces the direction d. Therefore I is equal to J when there is no constraint deleted. But, the index set I always satisfies $J_+ \subseteq I \subseteq J$.

Once the direction d is obtained, we then update the penalty parameter $\bar{\alpha}$ by $\bar{\alpha} = 10^\beta \alpha$, where β is the smallest nonnegative integer that satisfies

$$10^\beta \alpha \|A_I d\|_2 \geq \|u^{(I)}\|_2 \|(Ax-b)_+\|_2 . \qquad (2.5)$$

It is proved later that the penalty parameter remains bounded if problem (QP) is feasible, and that this method of choosing the penalty parameter makes d a downhill search direction of the penalty function P defined in (1.2).

A new point $\bar{x} = x + \bar{\lambda}d$ is then computed by doing a line-search on the function $P(\cdot, \alpha)$; that is,

$$P(x + \bar{\lambda}d, \bar{\alpha}) = \min_{\lambda \geq 0} P(x + \lambda d, \bar{\alpha}) .$$

We note here that the function $P(\cdot, \alpha)$ is strict convex and easy to evaluate; hence, there are no difficulties in the line-search computation.

3. BASIC PROPERTIES

In this section we study some basic properties of the
method. We first establish the result that, when problem
(QP) is feasible, a point x is its Karush-Kuhn-Tucker point
if and only if the direction d generated from the method is
zero. We need some lemmas.

Lemma 3.1: Let B and E be $p \times n$ and $q \times n$ matrices and
$\{h^k\}$ and $\{g^k\}$ be sequences of p-vectors and q-vectors, re-
spectively. If the system $By = h^k$, $Ey \leq g^k$ is consistent
for each k and if $h^k \to \bar{h}$ and $g^k \to \bar{g}$ then the system
$By = \bar{h}$, $Ey \leq \bar{g}$ is also consistent.

Proof: Notice that $By = h$, $Ey \leq g$ is consistent if and
only if

$$\begin{bmatrix} h \\ g \end{bmatrix} \in S + C$$

where S is the space spanned by the columns of the matrix
$\begin{bmatrix} B \\ E \end{bmatrix}$ and C is the closed convex cone generated by the col-
ums of $\begin{bmatrix} 0 \\ I \end{bmatrix}$.

By our assumption, for each k we have

$$\begin{bmatrix} h^k \\ g^k \end{bmatrix} \in S + C .$$

Because $S + C$ is a closed set and $h^k \to \bar{h}$ and $g^k \to \bar{g}$,
then we have

$$\begin{bmatrix} \bar{h} \\ \bar{g} \end{bmatrix} \in S + C ,$$

which implies our desired result.

<u>Lemma 3.2</u>: Let $Ay \leq b$ be consistent. Then there exists an $\bar{\varepsilon} > 0$ such that for any $x \in R^n$ and for any index set I with $J_+(x) \subseteq I \subseteq J(x)$, if $A_I y = b_I$ is inconsistent then $\|A_I(x^{(I)} - x)\|_2 \geq \bar{\varepsilon}$.

<u>Proof</u>: If the conclusion is not true, then for each positive integer k there exists some x^k and some index set I_k with $J_+(x^k) \subseteq I_k \subseteq J(x^k)$ such that

$$\lim_{k \to \infty} \|A_{I_k}(x^{(I_k)} - x^k)\|_2 = 0$$

but $r^k = A_{I_k} x^{(I_k)} - b_{I_k} \neq 0$. Since there are only a finite number of index sets, without loss of generality, we may assume that

$$J(x^k) = J, \quad J_+(x^k) = J_+, \quad I_k = I \quad \text{and} \quad r^k = r$$

for each k . Hence

$$r = A_I x^{(I)} - b_I = A_I(x^{(I)} - x^k) + A_I x^k - b_I .$$

Because $A_I(x^{(I)} - x^k) \to 0$, $J_+(x^k) \subseteq I$ and $A_I x^k - b_I \geq 0$, we have $r \neq 0$ and $r \geq 0$. Because $x^{(I)}$ is a least-squares solution of $A_I y = b_I$, we also have $A_I^T r = 0$.

Consider the system

$$\begin{cases} A_I y = b_I + r - A_I(x^{(I)} - x^k) \\ A_K y \leq b_K , \end{cases} \tag{3.1}$$

where $K = \{1, \ldots, m\} \backslash J$. The system (3.1) is consistent because x^k is a solution. It follows from Lemma 3.1 and $A_I(x^{(I)} - x^k) \to 0$ that the following system

$$\begin{cases} A_I y = b_I + r \\ A_K y \leq b_K \end{cases}$$

is also consistent and has a solution, say $\bar{y}$. Let $\theta(y) := \frac{1}{2}(Ay-b)_+^T(Ay-b)_+$. Then

$$\nabla\theta(\bar{y}) = A^T(A\bar{y}-b)_+ = A_I^T r = 0 .$$

It follows from the convexity of θ that $\bar{y}$ is a minimum point of θ and the minimum value of θ is $\theta(\bar{y}) = \frac{1}{2}r^T r \neq 0$. This contradicts the assumption that the system $Ay \leq b$ is consistent. Hence the proof is completed.

Theorem 3.3: Let quadratic program (QP) be feasible. If the direction d generated from the method is zero at a point x then x is a solution of problem (QP).

Proof: If $d = x^{(I)} - x = 0$, we want to show that $(x, u^{(I)})$ is a Karush-Kuhn-Tucker pair of problem (QP).

We first show that x is feasible. Because $A_I(x^{(I)}-x) = 0$ and $J_+ \subseteq I \subseteq J$, it follows from Lemma 3.2 that $A_I y = b_I$ must be consistent. Thus we have $A_I x^{(I)} = b_I$. Because $J_+ \subseteq I$ and $x = x^{(I)}$, it follows that $A_I x = b_I$, $J_+ = \phi$, and hence that x is feasible.

It follows from (2.2) that

$$Mx - c + A_I^T u_I^{(I)} = 0$$

$$A_I x = b_I .$$

We also have that $A_K x \leq b_K$, where $K = \{1,...m\}\backslash I$, and $u_i^{(I)} = 0$ for $i \in K$. Thus, we only need to show that $u_I^{(I)} \geq 0$.

From the method, there exists some $i \in I$ with $u_i^{(I)} < 0$ only if $x^{(I)} \neq x^{(J)}$. It follows from $A_J x^{(I)} = A_J x = b_J$ and $I \subseteq J$ that $x^{(I)}$ also solves the problem $\min_y \{ \frac{1}{2} y^T M y - c^T y \mid A_J y = b_J \}$. By the uniqueness of solution, we have $x^{(I)} = x^{(J)}$. This completes the proof.

When x solves problem (QP), it is obvious that the direction d generated at x is zero. Hence, we have the following corollary.

Corollary 3.4: Let quadratic program (QP) be feasible. A point x is a solution of problem (QP) if and only if the direction generated at x is zero.

Another straightforward consequence of Theorem 3.3 is contained in the following corollary, which is very useful in detecting infeasibility of the problem (QP).

Corollary 3.5: If x is not a Karush-Kuhn-Tucker point of problem (QP) and the direction d generated at x is zero then the quadratic program (QP) is infeasible.

We next show that, if problem (QP) is feasible, then the penalty parameter remains bounded.

Lemma 3.6: If problem (QP) is feasible, then there exists a positive number ρ such that for any $x \in R^n$ and for any index set I with $J_+(x) \subseteq I \subseteq J(x)$,

$$\| A_I (x^{(I)} - x) \|_2 \geq \rho \| (Ax - b)_+ \|_2 .$$

Proof: It follows from $A_I^T (A_I x^{(I)} - b_I) = 0$ that

$$\| (Ax - b)_+ \|_2^2 = \| A_I (x^{(I)} - x) \|_2^2 + \| A_I x^{(I)} - b_I \|_2^2 .$$

By Lemma 3.2 there exists $\varepsilon > 0$ such that $\|A_I(x^{(I)}-x)\|_2 \le \varepsilon$ implies $A_I x^{(I)} = b_I$. Hence, if $\|A_I(x^{(I)}-x)\|_2 \le \varepsilon$ then $\|A_I(x^{(I)}-x)\|_2 = \|(Ax-b)_+\|_2$. If $\|A_I(x^{(I)}-x)\|_2 > \varepsilon$, let

$$\eta := \max_K \{\|A_K x^{(K)} - b_K\|_2\}$$

where the maximum is taken over all index sets. Then

$$\|(Ax-b)_+\|_2^2 = \|A_I(x^{(I)}-x)\|_2^2 + \|A_I x^{(I)} - b_I\|_2^2$$

$$\le \|A_I(x^{(I)}-x)\|_2^2 + \eta^2$$

$$\le (1+\frac{\eta^2}{\varepsilon^2})\,\|A_I(x^{(I)}-x)\|_2^2 \ .$$

Therefore, the lemma follows if we set

$$\rho := \varepsilon/(\varepsilon^2+\eta^2)^{\frac{1}{2}} \ .$$

<u>Theorem 3.7</u>: If problem (QP) is feasible then the penalty parameter α in the method remains bounded.

<u>Proof</u>: By Lemma 3.6, $\|A_I d\|_2 = 0$ implies $\|(Ax-b)_+\|_2 = 0$. Therefore, we only need to show that when $\|A_I d\|_2 \ne 0$ the quantity

$$\|u^{(I)}\|_2 \cdot \frac{\|(Ax-b)_+\|_2}{\|A_I(x^{(I)}-x)\|_2}$$

remains bounded for any $x \in R^n$ and for any index set I with $J_+(x) \subseteq I \subseteq J(x)$.

The value $\|u^{(I)}\|_2$ is bounded because there are only a finite number of index sets and, as mentioned before, for

each index set I the vector $u^{(I)}$ is uniquely determined. By the previous lemma, we have that for some constant ρ

$$\frac{\|(Ax-b)_+\|_2}{\|A_I(x^{(I)}-x)\|_2} \leq \frac{1}{\rho} \, .$$

Hence, the proof is completed.

<u>Corollary 3.8</u>: If the penalty parameter $\alpha_k \to \infty$ in the method then problem (QP) is infeasible.

We observe that the penalty function $p(x,\alpha)$ is convex and for any direction d, its directional derivative $p'(x,\alpha;d)$ at x exists and is given by

$$p'(x,\alpha;d) = \begin{cases} d^TMx - c^Td + \dfrac{\alpha d^TA^T(Ax-b)_+}{\|(Ax-b)_+\|_2} \\ \qquad\qquad\qquad \text{if } x \text{ is infeasible;} \quad (3.2) \\[2ex] d^TMx - c^Td + \alpha\|(A_Jd)_+\|_2 \\ \qquad\qquad\qquad \text{if } x \text{ is feasible.} \end{cases}$$

The following two theorems show that the direction d generated from the method is a descent direction for the exact penalty function.

<u>Theorem 3.9</u>: Let quadratic program (QP) be feasible and let $x \in R^n$ be an infeasible point to (QP). If

$$\alpha \geq \|u^{(I)}\|_2 \frac{\|(Ax-b)_+\|_2}{\|A_Id\|_2} \, , \quad \text{then}$$

$$p'(x,\alpha;d) \leq -d^TMd \, .$$

<u>Proof</u>: Using (2.2.a) and (3.2) to eliminate c, we get

$$p'(x,\alpha;d) = -d^T M d - d^T A_I^T u_I^{(I)} + \alpha \frac{d^T A^T (Ax-b)_+}{\|(Ax-b)_+\|_2} \ . \qquad (3.3)$$

On the other hand, it follows from $J_+(x) \subseteq I \subseteq J(x)$ and $A_I^T(A_I x^{(I)} - b_I) = 0$ that

$$d^T A^T (Ax-b)_+ = d^T A_I^T (A_I x - b_I)$$

$$= d^T A_I^T (A_I x^{(I)} - b_I - A_I d)$$

$$= -d^T A_I^T A_I d$$

$$= -\|A_I d\|_2^2 \ .$$

Therefore, if

$$\alpha \geq \|u^{(I)}\|_2 \frac{\|(Ax-b)_+\|_2}{\|A_I d\|_2} \ ,$$

then

$$-d^T A_I^T u_I^{(I)} + \frac{\alpha d^T A^T (Ax-b)_+}{\|(Ax-b)_+\|_2} \leq \|A_I d\|_2 \|u^{(I)}\|_2 - \frac{\alpha \|A_I d\|_2^2}{\|(Ax-b)_+\|_2}$$

$$\leq 0 \ .$$

This, in conjunction with (3.3), implies our result.

Theorem 3.10: If x is a feasible point of problem (QP) and the vectors $\{a_i\}$, $i \in J(x)$, are linearly independent, then for any α,

$$p'(x,\alpha;d) \leq -d^T M d \ .$$

Proof: Using (2.2.a) and (3.2) to eliminate c, we have that

$$p'(x,\alpha;d) = -d^T M d - d^T A_I^T u_I^{(I)} + \alpha \|(A_J d)_+\|_2 \ .$$

We only need to show that $A_I d = 0$ and $a_j^T d \leq 0$ for any $j \in J \backslash I$, where $J = J(x)$.

Because x is feasible, we have $A_J x = b_J$. Hence, it follows from $I \subseteq J$ that the system $A_I y = b_I$ is consistent and $A_I x^{(I)} = b_I$. Therefore, we have that $A_I d = A_I(x^{(I)} - x) = 0$.

To show that $a_j^T d \leq 0$ for each $j \in J \backslash I$, we first show that the set $J \backslash I$ can not contain more than one indices if $\{a_i\}$, $i \in J$, are linearly independent. Assume that there are more than one indices in $J \backslash I$. Let q and p be the first and the second indices deleted from J, respectively. Then, by the method, the point $x^{(J)}$ is a Karush-Kuhn-Tucker point of the problem $\min \{\frac{1}{2} y^T M y - c^T y \mid A_K y = b_K,\ a_p^T y = b_p,\ a_q^T y = b_q\}$ and also of the problem $\min \{\frac{1}{2} y^T M y - c^T y \mid A_K y = b_K,\ a_p^T y = b_p\}$, where $K = J \backslash \{p, q\}$. Hence, there exists vectors u, v and numbers α, β and γ such that

$$Mx^{(J)} - c + A_K^T u + \alpha a_p + \beta a_q = 0$$

and

$$Mx^{(J)} - c + A_K^T v + \gamma a_p = 0 .$$

Here, we have $\beta < 0$ because, if not, the index q can not be deleted from J. Then, eliminating the term $Mx^{(J)} - c$ from the above two equations, we get

$$A_K^T(u-v) + (\alpha-\gamma)a_p + \beta a_q = 0 ,$$

which contradicts the assumption that $\{a_i\}$, $i \in J$, are linearly independent.

If $I = J$ then there is nothing to prove. Let j be the only index in $J \backslash I$. Then it follows from Theorem 1.9 in [6] that $a_j^T(x^{(I)} - x^{(J)}) < 0$. Hence we have that

$$a_j^T d = a_j^T(x^{(I)} - x)$$

$$= a_j^T(x^{(I)} - x^{(J)}) + a_j^T(x^{(J)} - x)$$

$$= a_j^T(x^{(I)} - x^{(J)}) \leq 0.$$

This completes our proof.

4. FINITE CONVERGENCE

In this section we show that, under suitable conditions, the method generates from any starting point a solution of Problem (QP) in a finite number of iterations.

The proof of finite convergence for the method is fairly long. To make it easier to understand, we outline our approach below. The proof is done by contradiction. If the method generates an infinite sequence $\{x^k\}$, then we show that the sequence enters from infeasible region into feasible region for only a finite times. This implies that, after finite iterations, either all the points will remain infeasible or all become feasible. Then we get a contradiction by showing that either case can not occur.

For convenience, we give the following definition.

<u>Definition 4.1</u>: Problem (QP) is said to be <u>regular</u> if it is feasible and for any feasible point x, the vectors $\{a_i\}$, $i \in J(x)$, are linearly independent.

We first present a lemma, which is a straightforward consequence of Theorem 3.7.

<u>Lemma 4.2</u>: If quadratic program (QP) is feasible then there exists a number α^* and an integer $\bar{k}$ such that the penalty parameter $\alpha_k = \alpha^*$ for all $k \geq \bar{k}$.

<u>Theorem 4.3</u>: If quadratic program (QP) is regular, then λ_k is greater than or equal to one for only a finite number of values of k.

<u>Proof</u>: By Lemma 4.1, after finite iterations and for a fixed α^*, the value of function $P(\cdot, \alpha^*)$ maintains strictly

monotone decreasing. If $d = x^{(I)} - x$, $\bar{x} = x + \bar{\lambda}d$ and $\bar{\lambda} \geq 1$ then at least one of the two inequalities

$$P(x, \alpha^*) > P(x^{(I)}, \alpha^*)$$

and

$$P(x^{(I)}, \alpha^*) > P(\bar{x}, \alpha^*)$$

is satisfied. Because there are only finite index sets I, the above inequalities can happen for only a finite number of times. Hence, the desired result follows.

Lemma 4.4: If x is infeasible and $J_+(x) \subseteq I \subseteq J(x)$ then for any λ in $[0,1)$, the point $(1-\lambda)x + \lambda x^{(I)}$ is infeasible.

Proof: If it is not true then there exists a λ in $[0,1)$ such that the point $\hat{x} = (1-\lambda)x + \lambda x^{(I)}$ is feasible; that is, $a_i^T \hat{x} \leq b_i$ for each i. Clearly, λ can not be zero and $\hat{x} = x + \lambda d$.

For any i we have that

$$a_i^T d = \frac{1}{\lambda} a_i^T (\hat{x}-x) \leq \frac{1}{\lambda} (b_i - a_i^T x) .$$

Therefore, it follows that $a_i^T d < 0$ if $i \in J_+(x)$ and $a_i^T d \leq 0$ if $i \in I \backslash J_+(x)$. On the other hand, we have that

$$a_i^T x^{(I)} - b_i = a_i^T (\hat{x}+(1-\lambda)d) - b_i$$

$$\leq (1-\lambda) a_i^T d .$$

This, in conjunction with the fact that $J_+(x) \neq \phi$, implies that

$$d^T A_I^T (A_I x^{(I)} - b_I) = \sum_{i \in I} (a_i^T d)(a_i^T x^{(I)} - b_i)$$

$$\geq \sum_{i \in J_+} (1-\lambda)(a_i^T d)^2 > 0 .$$

But, this inequality contradicts that $A_I^T(A_I x^{(I)} - b_I) = 0$. This completes our proof.

It follows from Lemma 4.4 that if x^k is infeasible and x^{k+1} is feasible then the stepsize λ_k must be no less than one. This result, in conjunction with Theorem 4.3, implies the following lemma.

Lemma 4.5: If quadratic program (QP) is regular and $\{x^k\}$ is a sequence of points generated by the method, there are only a finite number of consecutive pairs (x^k, x^{k+1}) in the sequence with x^k infeasible and x^{k+1} feasible.

Theorem 4.6: Let quadratic program (QP) be regular and $\{x^k\}$ be a sequence of points generated by the method, then one and only one of the following statements holds:

(a) The sequence $\{x^k\}$ terminates at a solution of (QP) in a finite number of steps;

(b) There exists $\bar{k}$ such that x^k is infeasible for all $k \geq \bar{k}$;

(c) There exists $\bar{k}$ such that x^k is feasible for all $k \geq \bar{k}$.

In the rest of this section we establish a finite convergence theorem for our method by showing that statements (b) and (c) in Theorem 4.6 actually can not hold.

We first deduce a contradiction when Statement (b) in Theorem 4.6 is true. Our deduction depends on that the

condition $A_I x^{(I)} = b_I$ always holds after a certain iteration. This, in turn, is derived from that the condition holds when the corresponding point x is in a slightly enlarged feasible region

$$X_\eta := \{y \,|\, Ay \le b + \eta e\} \qquad (4.1)$$

where e is the vector of ones and η is a suitable positive number. To determine the number η we first define

$$\delta^{(I)} := \max \{a_i^T x^{(I)} - b_i \,|\, i \in I\}$$

$$\delta := \min \{\delta^{(I)} \,|\, I \in \Omega\}$$

where

$$\Omega := \{I \,|\, A_I x^{(I)} \ne b_I \text{ and the system}$$

$$A_I y \ge b_I \text{ is consistent}\} .$$

Here, we use the convention that $\delta = \infty$ when $\Omega = \phi$. Then the number η in (4.1) is determined by

$$\eta := \frac{\delta}{\sqrt{2m}} . \qquad (4.2)$$

We need to show that η is positive, which follows from the following lemma.

__Lemma 4.7:__ If $I \in \Omega$ then $\delta^{(I)} > 0$.

__Proof:__ Let $r := A_I x^{(I)} - b_I$. If $I \in \Omega$ and $\delta^{(I)} \le 0$, then $r \ne 0$ and $r \le 0$. Let $\hat{x}$ be any point satisfying the system $A_I y \ge b_I$. Then

$$A_I(\hat{x} - x^{(I)}) \ge -r .$$

Premultiplying the above inequality by the nonzero and non-negative vector $-r$, we get

$$-r^T A_I(\hat{x}-x^{(I)}) \geq r^T r > 0 .$$

This contradicts that $A_I^T r = 0$, and hence the lemma is proved.

Corollary 4.8: The number η defined in (4.2) is positive.

Theorem 4.9: If $x \in X_\eta$ and $J_+(x) \subseteq I \subseteq J(x)$ then $A_I x^{(I)} = b_I$.

Proof: If it is not true then $I \in \Omega$ and

$$(A_I x^{(I)} - b_I)^T (A_I x^{(I)} - b_I) \geq \delta^2 .$$

It follows from $x \in X_\eta$ that

$$(A_I x - b_I)^T (A_I x - b) \leq m\eta^2$$
$$= \frac{\delta^2}{2} .$$

This contradicts that $x^{(I)}$ is a least-squares solution of $A_I y = b_I$. Hence the theorem is proved.

We next show that there are at most a finite number of iterations in which the condition $A_I x^{(I)} = b_I$ does not hold.

Lemma 4.10: Let x be any point in R^n. If $J_+(x) \subseteq I \subseteq J(x)$ and $A_I x^{(I)} \neq b_I$ then for each $\lambda \in [0,1]$, $(1-\lambda)x + \lambda x^{(I)} \notin X_\eta$.

Proof: Suppose that it is not true and for some λ in $[0,1]$ the point $\hat{x} = (1-\lambda)x + \lambda x^{(I)}$ is in X_η. By Theorem 4.9 and the definition of η, the number λ must be in the open

interval $(0,1)$. Let $K := \{i \mid a_i^T \hat{x} \geq b_i, \ i \in I\}$ and let j be an index in I that satisfies

$$\delta^{(I)} = a_j^T x^{(I)} - b_j .$$

It follows from Lemma 4.7 that $a_j^T x^{(I)} - b_j > 0$, which implies that $j \in K$. Therefore, because $\hat{x} \in X_\eta$, we have that

$$\sum_{i \in K} (a_i^T \hat{x} - b_i)^2 \leq m\eta^2$$
$$= \frac{\delta^2}{2}$$
$$< (\delta^{(I)})^2$$
$$\leq \sum_{i \in K} (a_i^T x^{(I)} - b_i)^2 .$$

On the other hand, for $i \in I \backslash K$ we have that

$$a_i^T x^{(I)} - b_i \leq a_i^T \hat{x} - b_i < 0 .$$

Hence, it follows that

$$\sum_{i \in I \backslash K} (a_i^T \hat{x} - b_i)^2 \leq \sum_{i \in I \backslash K} (a_i^T x^{(I)} - b_i)^2 .$$

Therefore, we have that $\|A_I \hat{x} - b_i\|_2 < \|A_I x^{(I)} - b_1\|_2$. This contradicts that $x^{(I)}$ is a least-squares solution of $A_I y = b_I$ and hence the proof is completed.

<u>Theorem 4.11</u>: If problem (QP) is regular then there are at most a finite number of iterations in which $A_I x^{(I)} \neq b_I$.

<u>Proof</u>: Suppose that at a certain iteration $A_I x^{(I)} \neq b_I$, then by Lemma 4.10 we have that the whole line-segment $[x, x^{(I)}]$ is contained in the complement set X_η^C of X_η. Define

$$Y := \overline{(X_\eta^C)} \cap \{y \mid P(y, \alpha^*) \leq P(x^{\bar{k}}, \alpha^*)\}$$

where $\overline{(X_\eta^C)}$ is the closure of X_η^C and $\bar{k}$ is a positive

integer such that for each $k \geq \bar{k}$, $\alpha_k = \alpha^*$ and $\lambda_k < 1$.
The compact set Y is contained in the infeasible region, in
which P is continuously differentiable, hence P has uni-
formly continuous derivatives on Y. Thus it follows from
the proof of Theorem 14.2.7 in [5] that there exists a forcing
function σ such that

$$P(x,\alpha^*) - P(\bar{x},\alpha^*) \geq \sigma(\frac{|\nabla P(x, *)^T d|}{\|d\|_2}) . \qquad (4.5)$$

The above inequality holds whenever $A_I x^{(I)} \neq b_I$. The se-
quence $\{P(x^k,\alpha^*)\}$ is monotone decreasing and bounded below
and hence convergent. If $A_I x^{(I)} \neq b_I$ for an infinite number
of times then it follows from (4.5) and $-\nabla P(x,\alpha^*)^T \geq d^T M d$
that for any $\epsilon > 0$ there exists $k \geq \bar{k}$ such that

$$A_{I_k} x^{(I_k)} \neq b_{I_k}$$

$$\|d^k\|_2 \leq \epsilon .$$

But this is impossible because it contradicts the conclusion
of Lemma 3.2. Hence our proof is completed.

We are now ready to show that Statement (b) of Theorem
4.6 can not hold.

<u>Theorem 4.12</u>: If problem (QP) is regular and $\{x^k\}$ is an
infinite sequence generated by the method then there exists
no $\bar{k}$ such that x^k is infeasible for all $k \geq \bar{k}$.

<u>Proof</u>: Suppose the conclusion is not true and there exists a
positive integer $\bar{k}$ such that x^k is infeasible for each
$k \geq \bar{k}$. By Theorem 4.3 and 4.11 we can choose $\bar{k}$ such that in
any iteration with $k \geq \bar{k}$, we have $\bar{\lambda} < 1$ and $A_I x^{(I)} = b_I$.

We want to show that in any iteration with $k \geq \bar{k}$ the set $J_+(x)$ is a proper subset of $J_+(\bar{x})$. This certainly implies a contradiction because there are only a finite number of index sets.

Let $i \in J_+(x)$, then it follows from $a_i^T x > b_i$ and $a_i^T x^{(I)} = b_i$ that

$$a_i^T \bar{x} = (1-\bar{\lambda}) a_i^T x + \lambda a_i^T x^{(I)}$$
$$> b_i .$$

Therefore, we have $J_+(x) \subseteq J_+(\bar{x})$. To show $J_+(x) \neq J_+(\bar{x})$ we consider function $P_I(\cdot, \alpha^*)$ defined by

$$P_I(x, \alpha^*) := \tfrac{1}{2} x^T M x - c^T x + \alpha^* \|(A_I x - b_I)_+\|_2 .$$

It is convex and the function

$$\phi(\lambda) := P_I(x + \lambda d, \alpha^*)$$

decreases monotonically for $0 \leq \lambda \leq 1$. In order to prove the monotonicity one obtains from (3.2) and (2.2.a) the bound

$$\phi'(1-) = d^T M x^{(I)} - c^T d - \alpha^* \|(-A_I d)_+\|_2$$
$$= -d^T A_I^T u^{(I)} - \alpha^* \|A_I d\|_2$$
$$\leq (\|u^{(I)}\| - \alpha^*) \|A_I d\|_2 .$$

Because of the lower bound on α that is given in the statement of Theorem 3.9, and because the proof of Lemma 3.6 shows that $\|(Ax-b)_+\|_2$ is no less than $\|A_I d\|_2$, it follows that $\phi'(1-) \leq 0$, which establishes the monotonicity. Therefore, we deduce from the continuity of derivatives in the infeasible region that the value $P(\bar{x}, \alpha^*)$ is not the same as the value of $P_I(\bar{x}, \alpha^*)$. It follows that $J_+(x) \neq J_+(x)$ and hence the proof is completed.

We now show that Statement (c) in Theorem 4.6 also can not hold.

<u>Theorem 4.13</u>: If problem (QP) is regular and $\{x^k\}$ is an infinite sequence that is generated from the method, then there exists no $\bar{k}$ such that x^k is feasible for all $k \geq \bar{k}$.

<u>Proof</u>: Suppose that there exists a $\bar{k}$ such that for each $k \geq \bar{k}$, the point x^k is feasible. By Lemma 4.2 and Theorem 4.3, we may assume that $\alpha_k = \alpha^*$ and $\lambda_k < 1$ for all $k \geq \bar{k}$.

We will establish a contradiction by showing that for $k \geq \bar{k}$, each set in the sequence $\{J(x^k)\}$ is different. Let Card(J) denote the cardinal number of the index set J. We first show that for $k \geq \bar{k}$, the sequence of numbers $\{\text{Card}(J(x^k))\}$ is nondecreasing.

Let x be a feasible point and d be the direction generated at x and $d = x^{(I)} - x$ for some index set I. As in the proof of Theorem 3.10, it follows from the regularity of the problem (QP) there is at most one constraint deleted and we have that $I = J$ or $I = J\backslash\{j\}$ for some $j \in J$. It follows that $A_J x = b_J$ and $I \subseteq J$ that the system $A_I y = b_I$ is consistent and $A_I x^{(I)} = b_I$. Therefore, if $\bar{x} = x + \bar{\lambda}d$ and $\bar{\lambda} < 1$ then at least a new constraint is added to I and we have that

$$\text{Card}(J(\bar{x})) \geq \text{Card}(J(x)) + 1 \quad \text{if} \quad I = J$$

$$\text{Card}(J(\bar{x})) \geq \text{Card}(J(x)) \quad \text{if} \quad I = J\backslash\{j\}.$$

This implies that for $k \geq \bar{k}$, the sequence of numbers $\{\text{Card}(J(x^k))\}$ is nondecreasing.

We now show that for $k \geq \bar{k}$, the sets $\{J(x^k)\}$ are all different. It is obvious that the index set $J(x)$ can not be repeated in the sequence $\{J(x^k)\}$ if $I = J$. This is

because that $\text{Card}(J(\bar{x})) \geq \text{Card}(J(x)) + 1$ and the sequence $\{\text{Card}(J(x^k))\}$ is nondecreasing. We only need to consider the case $I = J\backslash\{j\}$.

If $I = J\backslash\{j\}$, then by (2.4) we have that

$$\tfrac{1}{2}(x^{(J)})^T M x^{(J)} - c^T x^{(J)} \geq P(x^{(I)}, \alpha^*) \, .$$

By the line-search, we also have that

$$P(x^{(I)}, \alpha^*) \geq P(\bar{x}, \alpha^*) \, .$$

Let $\hat{x}$ be the point that solves

$$\min_y \{\tfrac{1}{2} y^T M y - c^T y \,|\, A_J y = b_J, \ A_K y \leq b_K\}$$

where $K = \{1,\ldots,m\}\backslash J$. Then,

$$P(\hat{x}, \alpha^*) = \tfrac{1}{2}\hat{x}^T M \hat{x} - c^T \hat{x} \geq \tfrac{1}{2}(x^{(J)})^T M x^{(J)} - c^T x^{(J)}$$

$$\geq P(x^{(I)}, \alpha^*)$$

$$\geq P(\bar{x}, \alpha^*) \, .$$

If z is any point that comes after the point x in the sequence $\{x^k\}$ with $J(z) = J(x)$, then it follows from the feasibility of z and $A_J z = b_J$ that

$$P(z, \alpha^*) = \tfrac{1}{2} z^T M z - c^T z \geq \tfrac{1}{2}\hat{x}^T M \hat{x} - c^T \hat{x}$$

$$\geq P(\bar{x}, \alpha^*) \, . \tag{4.6}$$

Because $z \neq \bar{x}$ and z comes after $\bar{x}$ in the sequence $\{x^k\}$, we have that $P(\bar{x}, \alpha^*) > P(z, \alpha^*)$, which contradicts (4.6). The proof is then completed.

We now conclude this section with the following finite convergence theorem.

Theorem 4.14: If quadratic program (QP) is regular then the method produces the solution of (QP) from any starting point in a finite number of steps.

Proof: This immediately follows from Theorems 4.6, 4.13 and 4.14.

5. COMPUTATIONAL RESULTS

Some preliminary computational tests have been carried out for the method in the Cyber 175 System at University of Illinois-Urbana. The test problems are randomly generated but positive definiteness of the matrix M is maintained. The starting point is chosen to be an unconstrained minimum point of the objective function. A linear system subroutine in LINPACK is used to solve the equality constrained problem (2.1), where the constraint $A_I^T A_I x = A_I^T b_I$ is taken care by a QR decomposition. A point is accepted as a solution when the Karush-Kuhn-Tucker conditions are satisfied within the tolerance 10^{-12}. A constraint is considered to be active if it is satisfied within 10^{-7}. We note here that the Cyber 175 has a 48 bit mantissa.

In the following tables, n is the dimension of variable vector and m is the number of inequality constraints. The numbers are the numbers of calls for a linear system subroutine, rather than those of iterations, which are usually smaller.

TABLE 1

M \ N	5	10	15	20	25	30	35	40	45	50
50	3	14	16	14	14	23	18	19	23	33
45	5	5	15	9	23	16	22	25	27	36
40	8	10	12	16	30	17	30	34	20	34
35	5	6	14	9	28	18	26	15	31	19
30	5	9	8	17	28	21	18	21	28*	17

*Indicates that the problem is infeasible.

TABLE 2

N \ M	5	10	15	20	25	30
25	4	6	3	11	32	25
20	3	8	9	10	15	9
15	5	7	12	12	40†	28
10	5	7	10*	19*	10	18*

*Indicates that the problem is infeasible.

†Indicates that the tolerance 10^{-12} has not been satisfied in 40 iterations.

According to these results, it seems that the number of iterations are usually less than the number of constraints and the efficiency of the method depends more on m than n. Therefore, the method is recommended for problems with fewer constraints.

The method becomes very efficient when we have a good but infeasible estimate of solution. This makes it useful in solving general nonlinear programming problems.

ACKNOWLEDGMENT

REFERENCES

[1]   R. Fletcher, "A general quadratic programming algorithm", Journal of the Institute of Mathematics and Its Applications, 7, 1971, 76-91.

[2]   P. E. Gill and W. Murray, "Numerical stable methods for quadratic programming", Mathematical Programming, 16, 1978, 349-372.

[3]   D. Goldfarb, "Extension of Newton's method and simplex method for solving quadratic programs", in Numerical Methods for Non-linear Optimization, Editor F. A. Lootsma, Academic Press, London, 1972.

[4]   S. P. Han and O. L. Mangasarian, "Exact penalty functions in nonlinear programming", Mathematical Programming, 17, 1979, 251-269.

[5]   J. M. Ortega and W. C. Rheinboldt, "Iterative solution of nonlinear equations in several variables", Academic Press, New York, 1970.

[6]   M. J. D. Powell, "Introduction to constrained optimization", Chapter 1 in Numerical Methods for Constrained Optimization, Editors P. E. Gill and W. Murray, Academic Press, London, 1974.

[7]   J. B. Rosen, "The gradient projection method for nonlinear programming, Part I. linear constraints", Journal Society of Industrial and Applied Mathematics, 9, 1961, 514-532.

REPORT DOCUMENTATION PAGE	READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
2180	AD-A099368	

4. TITLE (and Subtitle)

SOLVING QUADRATIC PROGRAMS BY AN EXACT PENALTY FUNCTION.

5. TYPE OF REPORT & PERIOD COVERED

Summary Report - no specific reporting period

6. PERFORMING ORG. REPORT NUMBER

7. AUTHOR(s)

Shih-Ping Han

8. CONTRACT OR GRANT NUMBER(s)

DAAG29-80-C-0041
NSF-ENG 79-03881

9. PERFORMING ORGANIZATION NAME AND ADDRESS

Mathematics Research Center, University of
610 Walnut Street   Wisconsin
Madison, Wisconsin 53706

10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS

Work Unit Number 5 -
Operations Research

11. CONTROLLING OFFICE NAME AND ADDRESS

(See Item 18 below)

12. REPORT DATE

February 1981

13. NUMBER OF PAGES

31

14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office)

15. SECURITY CLASS. (of this report)

UNCLASSIFIED

15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited.

MRC-TSR-2180

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Quadratic programming, gradient projection method,
exact penalty function

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

In this paper we study a gradient projection method for
quadratic programs that does not require the generated points
to be feasible and can avoid the computation of a feasible
starting point. This is done by using an exact penalty function
in the line-search. It is shown that the method can produce
from any starting point a solution in a finite number of
iterations.

DD FORM 1473 EDITION OF 1 NOV 65 IS OBSOLETE

UNCLASSIFIED